ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2014

Study of Mammogram Microcalcification to aid tumour detection using Naive Bayes Classifier

S.Krishnaveni[1], R.Bhanumathi[2], T.Pugazharasan[3]

Assistant Professor, Dept of CSE, Apollo Engineering College, Chennai, India[1]

Assistant Professor, Dept of CSE, Apollo Priyadarshanam Institute of Technology, Chennai, India [2]

PG student, ME (CSE), Apollo Engineering College, Chennai, India [3]

ABSTRACT: At present most of the women were having symptoms of breast cancer it can be detected by presence of microclacifications in mammogram. Classifier makes vital role in early detection and diagnosis of microclacifications in mammogram. Most of the classifier at present which is not more efficient to diagnosis the breast cancer. In this paper leads to analysis an efficient method by diagnosing the mammogram using Naive bayes classifier. The proposed method has a) ROI extraction (Chain code) b) Pre-processing (Enhancement), c) Feature extraction (HOG) and d) Classification using Naive bayes classifier. Naive bayes classifier is used to detect microcalcification at each location in the mammogram. It classifies the Mammogram images as Benign or Malignant. The test of the proposed system yield 96.5% microcalcification detection in mammograms. Experimental results show that the proposed method using Mammogram Image Analysis Society (MIAS) Database clinical mammogram.

KEYWORDS: Naive Bayes classifier, HOG, Chain code, Microcalcifications, Texture.

I.INTRODUCTION

For years cancer has been one of the biggest threats in human life, deaths caused by cancer are expected to increase in the future with an estimated 12 million people dying from cancer in 2030. Of all known cancers, breast cancer is a major concern among women. Treatment of breast cancer at an early stage can significantly improve the survival rate of patients. Mammography is currently the most sensitive method for detecting early breast cancer.Retrospective studies have shown that radiologists can miss the detection of a significant proportion of abnormalities in addition to having high rates of false positives. The estimated sensitivity of radiologists in breast cancer screening is only about 75%. In order to improve the accuracy of interpretation, a variety of Computer- Assisted Detection (CAD) techniques have been proposed.

In real sense, the Malignancy [15] or Benign, its type and from it detection of stage of cancer as invasive and non-invasive is a very fuzzy kind of decision making. Benign tumours are "well-differentiated," that the tumour cells differ only slightly in appearance and behaviour from their tissue of origin. Malignant or malignancy is used to describe a cancer that generally grows rapidly and is capable of spreading throughout the body. However, for the purpose of diagnostic analysis, classifications are suggested.

Detection of breast cancer is conducted by means of methods of mammography and ultra sono graphy (USG) imaging. The most frequent type of breast cancer, detected before the invasion stage, is ducal carcinoma in situ (DCIS). In this type of cancer, the most frequent markers are clusters of microcalcification. Microcalcifications clusters are one of the important radiographic indications related to breast cancer because they are present in 30%–50% of all cancers found mammographically. Imaging techniques play an important role in digital mammogram, especially of abnormal areas that naïve bayes classifier be felt but can be seen on a conventional mammogram. Before any image-processing algorithm of mammogram pre-processing steps are very important in order to limit the search for abnormalities without undue influence from background of the mammogram. These steps are needed only on digitized screen film mammography (SFM) images because digital mammography devices perform this step automatically during the image storing process. The segmentation process is easier on images obtained directly from the digital mammography devices.

Microcalcifications (MCCs) are tiny bits of calcium that may show up in clusters or in patterns (like circles) and are associated with extra cell activity in breast tissue. Scattered micro-calcifications are usually a sign of benign breast tissue. MCCs appear as small bright arbitrarily shaped regions on the large variety of breast texture background and characterize early breast cancer are detectable in mammograms shown in Fig1.For MCCs, the interpretations of their presence are very difficult because of its morphological features. For example, the sizes of MCCs are very tiny, typically in the range of 0.1mm- 1.0mm and the average is about 0.3mm. The dense tissues especially in younger women may easily be misinterpreted as MCCs due to film emulsion error, digitization artefacts or anatomical structures such as fibrous strands, breast borders or hypertrophied lobules that almost similar to MCCs. Other factors that contribute to the difficulty of MCCs detection are due to their fuzzy nature, low contrast and low distinguish ability from their surroundings.

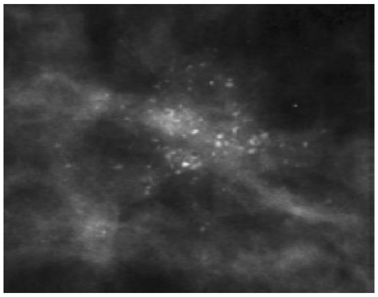


Fig1: Mammograms with Microcalcifications

However, microcalcification detection from mammograms may be troublesome. To overcome this problem, CAD is developed to improve the diagnostic accuracy and the consistency of the radiologists' image interpretation. Issam El-Naqaet al [4] proposed that Support Vector Machine Classifier used to automatically detect the presence of MCCs in a mammogram using two class pattern classifications to locate the position of the mammogram. Tomasz Arodza et al [5] used CAD system technique, in which filter the original image with microcalcification contrast shape, enhance by wavelet-based sharpening algorithm and visual analysis.The evaluated system significantly improves the detection of microcalcifications in small field digital mammography.Alain Tiedeu et al [1] proposed that clustered microcalcifications are detected on mammograms based on texture analysis. Ryohei Nakayama et al [2] proposed a Computer – Aided Diagnosis scheme using a Filter bank technique with Hessian Matrix to classifying NC (nodular component) and NLC (nodular and linear component). Alolfe et al [3] Computer aided diagnostic system based on wavelet analysis is proposed. Computer-Aided Diagnosis system that can be very helpful in diagnosing microcalcifications' patterns in digitized mammograms earlier and faster than typical screening programs.

Textures are one of the important features used for many applications. Texture features have been widely used in mammogram classification. The texture features are ability to distinguish between abnormal and normal cases. Texture can be characterized as the space distribution of gray levels in a neighbourhood [1, 2]. Texture feature have been proven to be useful in differentiating normal and abnormal pattern. Extracted texture features provide information about textural characteristics of the image. Different classifiers are used for medical imaging application including artificial intelligence, wavelet etc. Texture measures are two types, first order and second order. In the first order, texture measure are statistics calculated from an individual pixel and do not consider pixel neighbour relationships. Intensity feature are first order texture calculation. In the second order, measures consider the relationship between neighbour pixels GLCM is a second order texture calculation [16, 17]. Texture features has been extracted and used as parameter to enhance the classification result.

II. METHODOLOGY

A. Dataset Collection

It is difficult to access real medical images for experimentation due to privacy issue. To carry out experiments, data is collected from the Mammographic Image Analysis Society (MIAS). It consists of 322 images, which belong to three categories: normal, benign and malign, which are considered abnormal. In addition, the abnormal cases are further divided into six categories: circumscribed masses, speculated masses, Microcalcifications ill-defined masses, architectural distortion and asymmetry.

The framework of proposed approach is shown in Fig 2.

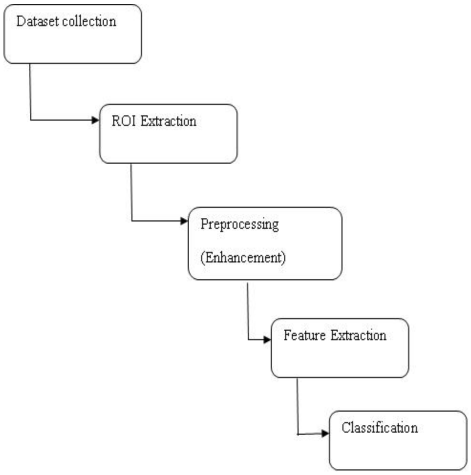
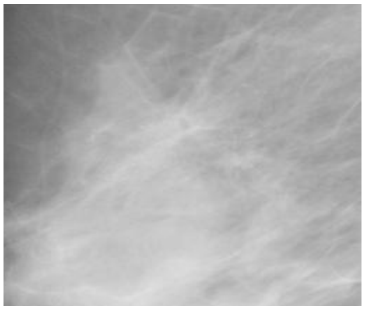


Fig 2: Framework of proposed approach

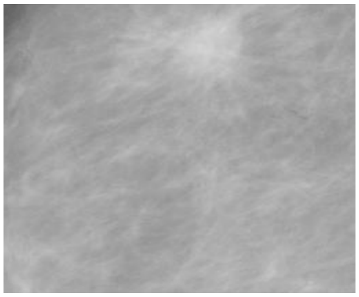
All images are digitized at a resolution of 1024×1024 pixels and eight-bit accuracy (gray level). The existing data in the collection consists of the location of the abnormality (like the centre of a circle surrounding the tumour), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumour type if exists (benign or malign). The proposed approach focus only on the microcalcification of benign and malignant images for abnormal and normal mammogram images.

B.ROI Extraction

The images taken from the MIAS databases are digitized at a resolution of 1024×1024 pixels and eight-bit accuracy (gray level). Since all the portion of mammogram images are not necessary for detection of microclassification of clusters, ROI extraction of certain portion is needed. Using the locations of any abnormalities supplied by the MIAS for each mammogram, the ROI of size 256×256 pixels is extracted by entering the coordinates of X, Y and radius in pixel. Then the images are divided into two sets the training set and the testing set respectively. The extracted portion of ROI of mammography with benign and malignant is shown in Fig 3.



(a)



(b)

Fig3: ROI extracted images (a) Benign b) Malignant

C. Preprocessing and Enhancement

Mammograms are medical images that are difficult to interpret, thus a pre-processing phase is needed in order to improve the image quality and make the segmentation results more accurate. The first step involves the removal of artefact and unwanted parts in the background of the mammogram. Then, an enhancement process is applied to the digital mammogram. Image enhancement operations can be used to improve the appearance of images, to eliminate noise or error, or to accentuate certain features in an image. The result of the process for detecting the suspicious lesions contains microcalcifications, and some noise. To reduce the amount of signatures which are not microcalcifications, perform a sequence of morphological operations on the filtered image.

The preprocessing and enhancement steps are as follows: Contrast Enhancement, Smoothening, Adaptive thresholding and top hat filtering. The original image is smoothed and subtracted from an image in which the contrast has been enhanced. The contrast enhancement technique increases the contrast of the image by mapping the values of the input intensity image to new values such that, by default, 1% of the data is saturated at low and high intensities of the input data. The Gaussian filters are used to smooth the original image, to attenuate as much as possible the MC signals while preserving at best the mammary tissue signals. The adaptive technique is used to isolate microcalcifications from their immediate surroundings rather than from a different region that may have a different density level. As long as the sizes of microcalcifications remain small and microcalcifications are not clustered too close to each other, this method will be able to preserve the original shapes of microcalcifications. The kernel size of 23 pixels is equivalent of approximately 1.2mm. Therefore, microcalcifications that are larger than 1.2mm in size are eliminated by the background correction process.

By subtracting pixel wise the smoothed-image from the contrast-enhanced image, the background (mammary tissue) is strongly attenuated. The difference-image $D(i,j)$obtained here above is binarized usinga local adaptive thresholding algorithm. For a rectangular window cantered on a pixel of coordinates(i,j) of width $w_{bin}$and heighth_{bin}, the pixel *(i, j)* in the resulting image is$BI(i,j)$ is computed as:

$$BI(i,j) = 1\, if \;\; if\; BI^{pix}(i,j) > \bar{m}_{i,j}^{bin} + k\sigma_{i,j}^{bin}$$

$$BI(i,j) = 0\; otherwise \qquad (1)$$

Where k is a pre-selected integer and the mean grey-level value $m_{i,j}^{bin}$and the standard deviation $\sigma_{i,j}^{bin}$are computed. A threshold is applied to the image such that only 2048 pixels of highest intensity (pixel value) are kept (assigned pixel value 1) and the rest pixels are assigned pixel values of 0. The number of pixels that are kept after thresholding is determined empirically that the sizes of microcalcifications in the threshold image are similar to those in the original image. Then it performs morphological top-hat filtering on the gray scale or binary input image. This filtering computes the morphological opening of the image (using `imopen`) and then subtracts the result from the original image. The transformed binarized image from ROI extracted image is shown in Fig 4.

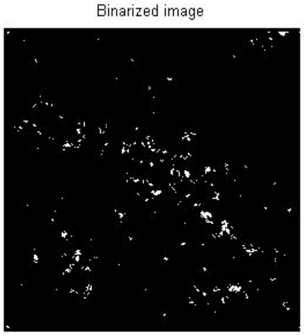


Fig 4: Transformed Binarized Image from ROI Extracted Image

D. Feature Extraction

In medical image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features. The features are used to extract the relevant information from the input data in order to perform the desired task.

1) **Histograms of Oriented Gradients**

Histogram of Oriented Gradients (HOG) is feature descriptors used in computer vision and image processing for the purpose of object detection. The technique counts occurrences of gradient orientation in localized portions of an image. This method is similar to that of histograms, scale descriptors, and shape contexts, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. The implementation of these descriptors can be achieved by dividing the image into small connected regions, called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then represents the descriptor. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination or shadowing. The preprocessed and enhanced mammography images features are extracted using Histograms of Oriented Gradients and its image and histogram is shown in Fig 5(a) and (b) respectively.

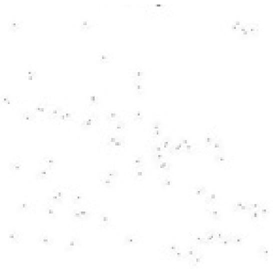


Fig 5 a): HOG image of Microcalcification of Mammogram

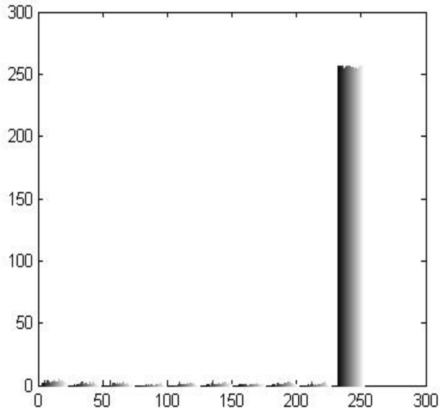


Fig 5 b) Histogram of HOG images

2) Texture Features:-

Intensity Based Features and GLCM based features are measured to analyse the performance. Intensity based features are first order statistics depends only on individual pixel values. The intensity and its variation inside the mammograms can be measured by features like mean and standard deviation using 40 samples of mammograms. The Gray Level Co-occurrence Matrix (GLCM) texture measurement is a method to analyse image texture [5, 6]. It is a robust method that has been developed for calculating first and second order texture features from image. It considers the relationship between two pixels at a time, the reference and neighbour pixel. The texture features are calculated as follow:

1) Mean Value:

The mean gives the average intensity value of an image. Mammographic images that contain micro calcifications have a higher mean than those of normal images. Mean calculated from the image as per the following equation.

$$\mu = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} P(i,j) (4)$$

Where 'i' *i*ndicates the rows of the image, 'j' indicates the columns of the image and P (i, j) is the cell denoted by the row and the column of the image.

2) Standard Deviation:

The standard deviation is a parameter closely associated with the mean. It refers to the dispersion of values in a mammographic image around the mean. Standard deviation is given as:

$$SD = \sqrt{(mean)^2} \qquad (5)$$

3) Energy:

Energy represents the orderliness of a mammographic image. Energy is generally given by the mean squared value of a signal. Energy calculated from the image as per the following equation.

$$E = \sum_{i,j=0}^{n-1} P(i,j)^2 \qquad (6)$$

Where 'i' indicates the rows of the GLCM matrix, indicates the columns of the GLCM matrix and *P(i, j)* is the cell denoted by the row and the column of the GLCM matrix.

4) Contrast: Contrast is a measure of the extent to which an object is distinguishable from its background. It represents the local variations present in an image, and calculates the intensity contrast between a pixel and its neighbour contrast calculated from the image as per the following equation.

$$C = \sum_{i,j=0}^{n-1} (i-j)^2 P(i,j) \qquad (7)$$

Where n denotes the number of pixels in the image and *P (i, j)* is the cell denoted by the row and column of the image.

5) Correlation: This measures the joint probability occurrence of the specified pixel pairs. It measures how correlated a pixel is to its neighbour over the whole image. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. So that it can be used to match the data to the predicted value.

$$Corr = \sum_{i,j=0}^{n-1} \frac{(i\times j)P(i,j) - \mu_i\mu_j}{\sigma_i\sigma_j} \qquad (8)$$

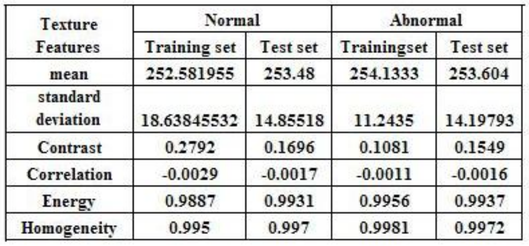
6) Homogeneity: It measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. These measure the values or pixel which is closer to the microcalcification of mammograms.

$$H = \sum_{i,j=0}^{n-1} \frac{P(i,j)}{1+|i-j|} \qquad (9)$$

The list of features used for experimental work for normal and abnormal images of mammograms is as shown in Table 1.

E. Classification

Naive Bayes Classifiers the process of learning to separate samples into different classes by finding common features between samples of known classes.



Texture Features	Normal		Abnormal	
	Training set	Test set	Trainingset	Test set
mean	252.581955	253.48	254.1333	253.604
standard deviation	18.63845532	14.85518	11.2435	14.19793
Contrast	0.2792	0.1696	0.1081	0.1549
Correlation	-0.0029	-0.0017	-0.0011	-0.0016
Energy	0.9887	0.9931	0.9956	0.9937
Homogeneity	0.995	0.997	0.9981	0.9972

Table 1: Texture feature measures for normal and abnormal of mammograms

It consist of input layer, hidden layer and output layer A set of samples may be taken from biopsies of two different tumour types, and their gene expression levels measured. Because making predictions on unknown samples is often used as a means of testing the Naive Bayes classifier. The two-layer feed-forward network is used. The training sets of images are compared with testing set of images and classify the tumour detected in those images. The classifier is used to classify the tumour as either normal or abnormal by detection of microcalcifications clusters of mammography images.

III. RESULT AND DISCUSSION

For implementation among the MIAS datasets of 322 images, fatty, dense and glandular tissue images of Normal(40) and Microcalcification of abnormal severity of benign or malignant (40) of 1024X1024 pixel images are considered. These images are normalized to 256X256 pixel of Region of Interest and stored as trained and testing set respectively. Pre-processing and enhancement is applied to remove of noise and artefact and improve the efficiency of the images. The Histogram of Oriented Gradients is applied and Gray Level Co-occurrence Matrix (GLCM) features and Intensity based features such as mean and standard deviations are measured. The classification is performed using Naive Bayes Classifier, in which network is created, trained for some samples and tested with remaining samples. The results are interpreted in confusion matrix is shown in Table. The confusion matrix describes actual and predicted classes of the proposed method.

True Positive (TP) – counts of all samples which are correctly called by the algorithm as being cancer.

False Positive (FP) – counts of all samples which are incorrectly called by the algorithm as being cancer while they are normal.

True Negative (TN) – counts of all samples which are correctly called by the algorithm as being normal.

False Negative (FN) – count of all samples which are incorrectly called by the algorithm as being normal while they are cancer.

A number of different measures are commonly used to evaluate the performance of the proposed method. These measures including Accuracy, sensitivity, specificity and precision are calculated from confusion matrix using the equation 10-13. It returns a value from -1(inverse prediction) to +1(perfect prediction).

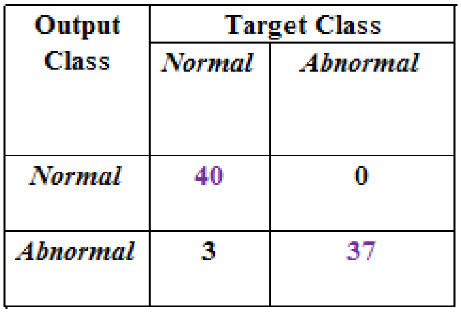
Accuracy = (TP+TN)/(TP+TN+FP+FN) (10)

Sensitivity = TP/(FP+TN) (11)

Specificity = TN/(FP+TN) (12)

Precision = TP/(TP+FP) (13)

The performance measure of proposed approach and comparison with other classifiers are formulated in Table 2. It shows that accuracy is 93.75%, specificity is 90%, sensitivity is 97.5% and precision is 97.2% for Naive Bayes Classifier. It reveals that better classification rate in accuracy and precision.



Output Class	Target Class	
	Normal	*Abnormal*
Normal	40	0
Abnormal	3	37

Table 2: Confusion Matrix for classification of breast cancer

Measures	NBC	SVM	k-NN
Accuracy	96.25%	93.75%	95%
Specification	92.50%	90%	90%
Sensitivity	100%	97.50%	100%
Precision	93.2%	90.70%	90.91%

Table3: Performance analysis of various classifiers

IV. CONCLUSION

Medical Imaging refers to view the human body in order to diagnose, monitor or treat medical conditions. Breast cancer has become a public health problem among women around the world. The features are extracted using HOG n provides statistical features. The texture features such as glcm features and mean and standard deviations are measure and the classification of mammograms using Naive Bayesclassifier is presented. The proposed experimental results show that when compared to several other methods Naive Bayes Classifier shows accuracy 96.25% microcalcification detection in mammograms. The high accuracy of classification of MCCs obtained with the proposed Naive Bayes classifier can help radiologists in making an accurate diagnostic decision, which can reduce unnecessary biopsies.

REFERENCES

[1] Alain Tiedeu, Christian Daul, "Texture-based analysis of clusteredmicrocalcifications detected on mammograms",Elsevier Journal of Digital Signal Processing, VolNo.22(1), January 2012, pp.124-132.

[2] Ryohei Nakayama, Yoshikazu Uchiyama, "Computer-Aided Diagnosis Scheme Using a Filter Bank for Detection of Microcalcification Clusters in Mammograms", IEEE Transactions On Biomedical Engineering, Vol. No.53(2), February 2006, pp.273-283.

[3] M. A. Alolfe, A. M. Youssef, "Computer-aided Diagnostic System Based On Wavelet Analysis ForMicrocalcification Detection In Digital Mammograms", IEEE Bio-Medical Engineering Conference, December, 2008.

[4] R.Nithya,B.Santhi,"comparative study of feature extraction, method for breast cancer classification", Journal of theoretical and applied Information Technology, vol.33(2),2011, pp 220-224.

[5] Tomasz Arodza, MarcinKurdziel, "Detection of clustered microcalcifications in small field digital mammography ",Elsevier Journal of Computer Methods and Programs in Biomedicine, 81, 2005,pp.56-65.

[6] J Suckling , S Astley, D Betal, N Cerneaz, D R Dance,"*The Mammographic Image Analysis Society Digital Mammogram Database*ExerptaMedica. International Congress Series 1069, 1994, pp375-378.

[7] Samir Kumar Bandyopadhyay,"Pre-processing of Mammogram Images", International Journal of Engineering Science and Technology, Vol. 2(11), 2010, pp.6753-6758.

[8] V. Saravanan, S.PitchumaniAngayarkNaivebayesclassifieri, "A Novel Approach for Cancer Detection in MRI Mammogram Using Decision Tree Induction and BPN",International Journal of Image Processing (IJIP), Vol4(6), pp.661-668.

[9] Vishnukumar K. Patel, Prof. Syed Uvaid, Prof. A. C. Suthar, " Mammogram of Breast Cancer detectionBased using Image Enhancement Algorithm", International Journal of Emerging Technology and Advanced Engineering, Vol.2(8), August 2012, pp. 143-147.

[10] Dheeba. J, WiselinJiji. G, "Detection of Microcalcification Clusters in Mammograms using Neural Network", International Journal of Advanced Science and Technology Vol.19, June, 2010, pp.13-22.

[11] Fatima Eddaoudi, FakhitaRegragui, "Microcalcifications Detection in Mammographic Images Using Texture Coding", Applied Mathematical Sciences, Vol. 5(8), 2011, pp. 381 – 393

[12] JawadNagi, Sameem Abdul Kareem, FarrukhNagi, Syed Khaleel Ahmed, "Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms", IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 30th November - 2nd December 2010, pp. 87-92.

[13] A. K. Jain, "Fundamentals of Digital image Processing", Englewood Cliffs, Prentice Hall, 1989.

[14] Rafael C Gonzalez, Richard E Woods, Steven L. Eddins, Digital Image Processing using MATLAB,2nd ed, Pearson Education, Singapore, 2005.

[15] W. liyang, Y. Yongyi, N. Robert, and J. Yulei, "A study on Several machine-Learning Methods for Classification of Malignant and Benign Clustered microcalcifications," *IEEE transactions on medical imaging*, vol. 24(3), March 2005.

[16] M.Vasantha, V. SubbiahBharathi, "Medical Image feature extraction selection and classification", International Journal of engineering science and technology vol.2 (6), 2010,pp. 2071-2076.